NASA

Technical Memorandum 78088

Information Theory and the Earth's Density Distribution

(NASA-TM-78088) INFORMATION THEORY AND THE EARTH'S DENSITY DISTRIBUTION (NASA) 26 p HC A03/MF A01 CSCL 08G

N78-20692

Unclas

G3/46 12570

David Parry Rubincam

FEBRUARY 1978

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

TM 78088

INFORMATION THEORY AND THE EARTH'S DENSITY DISTRIBUTION

David Parry Rubincam
NAS-NRC Resident Research Associate

February 1978

GODDARD SPACE FLIGHT CENTER
Greenbelt, Maryland

CONTENTS

LIST OF FIGURES

PRECEDING PAGE BLANK NOT FILMED

INFORMATION THEORY AND THE EARTH'S DENSITY DISTRIBUTION

David Parry Rubincam
NAS-NRC Resident Research Associate

ABSTRACT

The present paper argues for using the information theory approach of Jaynes (1957) as an inference technique in solid earth geophysics. A spherically symmetric density distribution is derived as an example of the method. A simple model of the earth plus knowledge of its mass and moment of inertia leads to a density distribution which is surprisingly close to the optimum distribution of Bullen (1975). Future directions for the information theory approach in solid earth geophysics as well as its strengths and weaknesses are discussed.

INFORMATION THEORY AND THE EARTH'S DENSITY DISTRIBUTION

1.0 INTRODUCTION

We wish to introduce to solid earth geophysics a method of scientific inference which has had great success in statistical mechanics (see, e.g., Jaynes (1957, 1963); Tribus (1961); Katz (1967); and Baierlein (1971)) and in spectral analysis (e.g. Burg (1972); Smylie et al. (1973); and Graber (1976)). It is the information theory approach of Jaynes (1957), based upon Shannon's (1948) information measure. We will illustrate the approach by inferring a density distribution for the earth based on knowledge of its mass and moment of inertia. The earth is assumed to be spherical and the density distribution spherically symmetric.

The nature of the inference problem is the following. We desire to know what the density distribution $\rho(r)$ is as a function of radial distance r from the center of the earth. Suppose the only information we have is its mass M_E and moment of inertia C_E, both of which depend upon $\rho(r)$. Clearly we do not have enough information to say what the density distribution $\rho(r)$ actually is. Any proposed distribution which satisfies the mass and moment of inertia is nonunique; there are infinitely many other distributions which also satisfy the given data.

There are several methods for dealing with this problem. (For a general discussion see Bullen (1975, pp. 60-64).) The approach of Backus and Gilbert (1967, 1968) is to study all solutions consistent with the given data; this is called the geophysical inverse problem. The Backus-Gilbert approach has been used extensively. See, for example, Gilbert et al. (1973); Parker (1977a, 1977b); Jordan and Franklin (1971); and references cited by Parker (1977a,

1977b), Richards (1975), Anderson (1975), and Engdahl et al. (1975). Press (1968a, 1968b) adopted a Monte Carlo technique of testing a wide range of models against the data and retaining only those which agreed with it. However, the commonest method by far is modeling: by introducing other assumptions, the answer becomes unique. The assumption of the Adams-Williamson equation, for instance, plus the known mass, moment of inertia, seismic velocities, and surface density determine a unique density distribution (Alterman et al. 1959, pp. 80-81). Of course a difficulty with this approach is that the assumed conditions may not hold.

Suppose we look at the problem from the following viewpoint. If we had to pick one answer (in our case, density distribution) from all the possible answers which fit the data, which would we pick as the most likely? To put it bluntly, what is our "best guess"? It is of extreme interest that the information theory approach of Jaynes (1957) provides an answer to this question. (Baierlein (1971) has an excellent general discussion of the information theory approach.)

2.0 INFORMATION MEASURE

At the heart of the approach is Shannon's (1948) information measure

$$MI(P_1, P_2, \ldots P_N) = -K\sum_{i=1}^{N} P_i \ln P_i \tag{2.1}$$

Here P_i is the probability that the ith of N possible answers is true and K is a positive constant. This function was originally termed the entropy function (Tribus and McIrvine, 1971, p. 180), due to its similarity to thermodynamic entropy. For this reason the information theory approach is often called the maximum entropy method, or MEM for short. The relationship between the information measure and thermodynamic entropy is deep, but the

two are not identical (Baierlein, 1971, pp. 473-478). To avoid confusion we will follow Baierlein (1971, p. 64) and call Shannon's information measure $MI(P_1, P_2, \ldots, P_N)$, where MI stands for "Missing Information", or the amount of information needed to determine which answer is correct.

We will not prove that MI is a measure of missing information; proofs are given by Shannon (1948) and Baierlein (1971). Rather, we will merely indicate its plausibility with an example. But first we note that $MI \geq 0$; the amount of information needed to single out the correct answer is never negative. This is certainly an intuitively desirable property. Now let us suppose that all of the probabilities are equal. In this case MI attains its maximum value. This accords with intuition — we are surely in a state of maximum ignorance (i.e. need the most information) if we can favor no answer above another in terms of probability. Suppose now we have discovered that the ith possibility is the correct answer. Then $P_i = 1$ and $P_j = 0$ for $j \neq i$. How much information is missing now? In this case $P_i \ln P_i = 1 \ln 1 = 0$, and $P_j \ln P_j = 0$ for $j \neq i$ (by virtue of $\lim_{x \to 0} x \ln x = 0$). Thus $MI = 0$; no information is missing, we have the answer. This also accords with intuition. Normally our ignorance lies between these two extremes, and MI takes on values accordingly between its maximum and 0. Hence MI is a plausible measure of missing information.

We should point out here that MI is not dimensionless (Edmundson, private communication, 1976), a fact that does not appear to be explicitly noted in Tribus (1961), Katz (1967), or Baierlein (1971). It carries units of information. For example, if we change the base of the logarithm in eq. (2.1) from e to 2, which changes K to a new constant K', and set $K' = 1$, then $MI = -\Sigma P_i \log_2 P_i$ and MI is measured in bits. In the following development we will retain the natural logarithm base and set $K = 1$, so that MI is measured in

nats (from natural units). We will suppress the units in the following development, but it should be remembered that MI is not a dimensionless quantity.

3.0 JAYNES' PRINCIPLE OF MINIMUM PREJUDICE

The essence of the information theory approach is this: choose the probabilities P_1, P_2, ... P_N of the possible outcomes to make MI as large as possible, subject to the constraints of the known data. This is Jaynes' principle of minimum prejudice (Tribus and Rossi, 1973). Hence the information theory approach is a rational method for assigning probabilities.

Let us illustrate the technique with an example. Suppose that we do not know the mass of the earth exactly, but (due to experimental error, say) it must be chosen from the values M_1, M_2, . . . M_N. Aside from $\Sigma P_i = 1$, this is all we know. We must find P_i, the probability that M_i is the correct mass, by maximizing MI. This is done by taking the partial derivative of

$$-\sum_{i=1}^{N} P_i \ln P_i + \alpha_0 \sum_{i=1}^{N} P_i$$

with respect to each P_i and setting it equal to zero. The α_0 is a Lagrange multiplier which insures that all of the probabilities add up to 1. Carrying out the process yields

$$-\ln P_i - 1 + \alpha_0 = 0$$

or

$$P_i = e^{\alpha_0 - 1} = \text{constant}$$

The unknown α_0 may be found from the constraint

$$\sum_{i=1}^{N} P_i = 1$$

giving

$$P_i = 1/N$$

All of the probabilities are equal. We know nothing about the various M_i to favor one particular value over another.

Now suppose we obtain further information: we learn that the expectation value of the mass is

$$\sum_{i=1}^{N} P_i M_i = \overline{M}_F$$

We reassign probabilities in accordance with Jaynes' principle:

$$\frac{\partial[-\Sigma P_i \ln P_i + \alpha_0 \Sigma P_i + \alpha_1 \Sigma P_i M_i]}{\partial P_i} = 0 \qquad i = 1, 2, \ldots N$$

giving

$$P_i = e^{\alpha_0 - 1} e^{\alpha_1 M_i}$$

where α_0 and α_1 are Lagrange multipliers to be found from the constraints

$$\Sigma P_i = 1, \ \ \Sigma P_i M_i = \overline{M}_F .$$

Note that our method is completely analogous to that of the canonical ensemble in statistical mechanics. Indeed, the mathematics are identical. The only difference is in the philosophical basis, which indicates that the method has broad applicability and is not confined only to statistical mechanics.

Obviously assigning probabilities is rather easy. But clearly the probabilities do not represent frequencies; their values change the moment we acquire new information. What is the point in maximizing MI? Why not do something else?

The users of Jaynes' principle have a powerful argument in its favor. The probabilities are indeed not mere frequencies, they say. A probability represents the "degree of

rational belief" (Baierlein, 1971, p. 13) that a particular answer is correct, a more general notion than a frequency. (See Cox (1946, 1961) for the quantitative basis for this view of probability.) And since MI measures the amount of information needed to determine the correct answer, **any method for assigning probabilities which does not maximize MI under known constraints (knowledge) tacitly assumes information it hasn't got!** In other words, if someone assigns probabilities not in accordance with Jaynes' principle, that person is prejudicing the probabilities without foundation in the known data. Thus the name, "principle of minimum prejudice."

This point is particularly clear in the example where we knew one of the M_i was the correct answer, but had no other information (other than $\Sigma P_i = 1$). In this case Jaynes' principle assigns equal probabilties to all outcomes. We are completely ignorant as to which answer is correct. If someone uses some other principle, and assigns (say) a larger probability to M_1 than to the other M_i, we can say, "You favored M_1 as being the most likely mass over all of the others. What basis (i.e. information) do you have to do that?" While the argument is powerful, the information theory approach is not without its problems. We will discuss some of these later.

4.0 INFORMATION THEORY DENSITY DISTRIBUTION

We are almost in a position to find the information theory density distribution inside the earth, using knowledge of the expectation values of the mass and moment of inertia. The only thing left to do is set up the problem. We will make heavy use of the methods of statistical mechanics; particularly, the grand canonical ensemble.

Imagine a three-dimentional Cartesian coordinate system with its origin at the center of the earth. The grid system will divide up the earth into many cubes of identical

volume V, just as ordinary graph paper divides up a plane into squares of equal area. We can approximate the spherical surface of the earth as closely as we like by making the cubes as small as we like. Let $\vec{r}_i$ be the vector from the center of the earth to the jth cube and set $|\vec{r}_i| = r_i$. Let the mass of the earth be the sum of the masses of a large number of indistinguishable particles, each with mass m. Let there be n_i particles in the ith cube. The mass M_E and moment of inertia C_E of the earth are then

$$\begin{aligned} M_E &= \sum_i n_i m \\ C_E &= 2/3 \sum_i n_i m r_i^2 \end{aligned} \tag{4.1}$$

where the subscript i runs over all the cubes comprising the earth. The factor (2/3) in the second equation makes use of the fact that the density distribution is spherically symmetric, and takes care of r_i being the distance from the center of the earth to a cube and not the distance to some axis of rotation.

Let us remark here that we have chosen cubes of equal volume so as to treat all regions of the earth identically, and indistinguishable particles because the interchanging of particles leaves the density distribution unaffected. We make no commitment as to the values of m and V. As we shall see, they drop out of the final equation for the density distribution.

We are ready to begin. Our information will be that the earth may be made up of any number of particles, but that the expectation value of the mass $\Sigma P_i M_i$ and moment of inertia $\Sigma P_i C_i$ are known to be $\bar{M}_E$ and $\bar{C}_E$, respectively. In practice, $\bar{M}_E$ and $\bar{C}_E$ will be the experimentally determined values. What we will do is find $\bar{n}_i$, the expectation value for the number of particles in the ith cube. The probabilities are computed according to Jaynes' principle of minimum prejudice:

$$\frac{\partial[-\Sigma P_i \ln P_i + \alpha_0 \Sigma P_i + \alpha_1 \Sigma P_i M_i + \alpha_2 \Sigma P_i C_i]}{\partial P_i} = 0$$

giving

$$P_i = \frac{e^{\alpha_1 M_i + \alpha_2 M_2}}{Z}$$

where

$$Z = e^{1-\alpha_0} = \sum_i e^{\alpha_1 M_i + \alpha_2 C_i} \qquad (4.2)$$

From eq. (4.1) we may write

$$\frac{M_i}{m} = \sum_i n_j$$

$$\frac{3C_i}{2m} = \sum_i n_j r_i^2 \qquad (4.3)$$

where the subscript i on the n_j has been suppressed. The problem now looks exactly like that of the grand canonical ensemble, with n_j playing the role of occupation numbers, r_i^2 the role of energy levels, and eq. (4.2) the grand partition function. The treatment of this problem may be found in any standard statistical mechanics text. We choose to follow Morse (1969, pp. 322-326).

Using eq. (4.3) in eq. (4.2), we have

$$Z = \sum_i e^{\alpha_1 \sum_i n_i + \alpha_2 \sum_i n_i r_i^2} \qquad (4.4)$$

where we have redefined $\alpha_1 m$ as α_1 and $2/3\alpha_2 m$ as α_2. Note that

$$\frac{\partial \ln Z}{\partial \alpha_1} = \sum_i \left(\sum_i n_j\right) \frac{e^{\alpha_1 \sum_i n_i + \alpha_2 \sum_i n_i r_i^2}}{Z} \qquad (4.5)$$

$$= \sum_i n_j \sum_i \frac{e^{\alpha_1 \sum_i n_i + \alpha_2 \sum_i n_j r_i^2}}{Z} = \sum_i \bar{n}_i$$

a result that we will make use of shortly.

Let us now rewrite eq. (4.4) as a summation over the possible values of n_j instead of over i. For distinguishable particles it is (Morse, 1969, p. 324)

$$Z = \sum_{n_1, n_2 \ldots} \frac{N!}{n_1!\, n_2! \ldots} e^{\alpha_1 \sum_j n_j + \alpha_2 \sum_j n_j r_j^2}$$

with $N = n_1 + n_2 + \ldots$.

The thing to do now is make Z mathematically tractable. We do this in the following manner. Assume the cubes are so small that the chances of two particles sharing the same cube are negligible. This is equivalent to assuming the particles follow Maxwell-Boltzmann statistics. To take care of the indistinguishability of the particles we can then divide the above equation by N! and obtain

$$Z = \sum_{n_1, n_2 \ldots} \frac{1}{n_1!\, n_2!} e^{\alpha_1 \sum_j n_j + \alpha_2 \sum_j n_j r_j^2}$$

as the approximate value for the grand partition function. Further, since

$$1 \cong \frac{1}{n_1!\, n_2! \ldots}$$

we can separate Z into factors for each cube:

$$Z = \sum_{n_1} e^{(\alpha_1 + \alpha_2 r_1^2) n_1} \cdot \sum_{n_2} e^{(\alpha_1 + \alpha_2 r_2^2) n_2} \ldots$$
$$= Z_1 \cdot Z_2 \cdot \ldots$$

where

$$Z_j = \exp(e^{\alpha_1 + \alpha_2 r_j^2})$$

by virtue of

$$e^x = \sum_{n=0}^{\infty} x^n / n!$$

From eq. (4.5), and above we have

$$\frac{\partial \ln Z}{\partial \alpha_1} = \sum_j e^{\alpha_1} e^{\alpha_2 r_j^2} = \sum_j \bar{n}_j$$

where evidently

$$\bar{n}_j = e^{\alpha_1} e^{\alpha_2 r_j^2}$$

The density distribution is obviously

$$\bar{\rho}(\vec{r}_j) = \frac{m}{V} e^{\alpha_1} e^{\alpha_2 r_j^2} \quad .$$

By the assumption of spherical symmetry for the density distribution we can drop the subscript and write

$$\bar{\rho}(r) = \bar{\rho}(0) e^{\alpha_2 r^2} \tag{4.6}$$

which we will take as the desired information theory density distribution. The two constants $\bar{\rho}(0) = \frac{m}{V} e^{\alpha_1}$ and α_2 may be found from our knowledge of the expectation value for the mass and moment of inertia:

$$\bar{M}_E = 4\pi \int_0^{a_E} \bar{\rho}(r) r^2 \, dr = 5.976 \times 10^{27} \text{ gm} \tag{4.7}$$

$$\bar{C}_E = \frac{8\pi}{3} \int_0^{a_E} \bar{\rho}(r) r^4 \, dr = 8.068 \times 10^{44} \text{ gm cm}^2 \tag{4.8}$$

where a_E is the radius of the earth and our numerical values have come from Stacey (1969, p. 279). In eq.s (4.7) and (4.8) we have assumed that the cubes are so small that we may switch from summations to integrals without serious error. By numerical integration of eq.s (4.7) and (4.8), we find that

$$\bar{\rho}(r) = 12.30 e^{-1.46 r^2 / a_E^2} \text{ gm/cm}^3 \tag{4.9}$$

is our "best guess" for the density distribution based on the given data.

A plot of eq. (4.9) appears in Fig. 1, along with the "optimum" density distribution given by Bullen (1975, p. 361), which presumably gives the most plausible distribution on the basis of all the known data. The two curves agree remarkably well, in view of the fact that the information theory density distribution makes use of only two basic pieces

of data: mass and moment of inertia. No seismic data or equations of state have been included in our information.

5.0 FUTURE DIRECTIONS FOR THE THEORY

The result obtained above may be easily generalized to include any known volume integrals of the density distribution. Supposing that there are I such integrals having the form

$$\int_{\substack{\text{volume}\\ \text{of earth}}} \rho(\vec{r}) f_i(\vec{r}) dv = F_i \qquad (i = 1, \ldots, I)$$

the resulting average density distribution is

$$\rho(\vec{r}) = \text{const} \cdot \exp(\alpha_1 f_1(\vec{r}) + \alpha_2 f_2(\vec{r}) + \ldots + \alpha_I f_I(\vec{r}))$$

The Lagrange multipliers α_i are found from the known values F_i. Note that the above result is not restricted to the spherically symmetric case. Besides the mass and moment of inertia, the spherical harmonic coefficients $\overline{C}_{lm}$ and $\overline{S}_{lm}$ of the earth's gravitational potential immediately come to mind as integrals having this form. We intend to publish the resulting $\rho(\vec{r})$ based on the gravity field coefficients in the near future.

The next obvious extension of the theory is to assume the earth is an elastic body so as to include the elastic parameters $\mu(\vec{r})$ and $\lambda(\vec{r})$ in addition to the density distribution $\rho(\vec{r})$ as unknown quantities to estimate. This will allow seismic travel times, free oscillation periods, and body tide observations to used, all of which depend upon $\mu(\vec{r})$, $\lambda(\vec{r})$ and $\rho(\vec{r})$. Graber (1977) has made a start in this direction using torsional vibrational modes of the earth, but much work needs to be done. Inclusion of chemical composition and equations of state are other possible directions for the theory.

The goal here is to put into the problem all of the physics and data that we know and maximize the remaining missing information. If we come to a point where the physics

and data determine a unique answer, there is no need for information theory or any other inference technique. There is no remaining uncertainty; we have the answer. Since this happy state of affairs is unlikely to occur soon, geophysics has a real need for sound methods of inference.

6.0 DISCUSSION

We will discuss here some problems and questions about the information theory approach. The two main problems which face it have been discussed extensively in the literature and only bear indirectly on the geophysical problem at hand. Hence we will merely point them out and then move on to questions more relevant to geophysics.

The first problem is that the information theory approach appears unable to deal with certain kinds of information. For example, suppose we flip a coin 100 times and it comes up heads 75 times. Clearly we have information on whether the next flip will be a heads or a tails. But how do we maximize eq. (2.1) using this data? Information theory seems to be silent on this question. This coin flip problem is discussed by Rowlinson (1970) and references he cites.

The second question is what to do about continuous probability distributions. It has vexed even the most ardent proponents of the information theory approach.

Shannon (1948, p. 628) proposed as the appropriate generalization for the continuous case in one dimension the function

$$MI = -K \int_{-\infty}^{\infty} p(x)\ln p(x)dx$$

where $p(x)$ is the probability distribution and x is a continuous parameter. One difficulty is immediately apparent. It takes the logarithm of a dimensioned quantity: $p(x)$ has dimensions of inverse length if x is a length, say. The biggest stumbling block, however, to the use of the above equation is in assigning prior probabilities. To illustrate, suppose

x represents the speed of a particle and is known to lie between values x_1 and x_2. We have no further information. If we apply Jaynes' principle we find $p(x)$ = constant for $x_1 \leq x \leq x_2$ and zero outside the interval. But if we had taken the kinetic energy of the particle (which depends on x^2) as the continuous parameter and found the probability distribution as a function of kinetic energy, we would have again found it to be constant over the interval. The two are inconsistent: a constant distribution for speed implies a nonconstant distribution for energy and vice versa. Thus the problem. Jaynes (1963, 1968) and Hobson and Cheng (1973) argue that the above equation must be modified, while Tribus and Rossi (1973) and Batty (1974) feel that it is the correct equation and that the inconsistency is a pseudoproblem. Rowlinson (1970) also discusses this matter.

Getting back to the density distribution derived here, a question arises: why assume that the cubes are sparsely occupied, thus giving Maxwell-Boltzmann statistics? This would seem all right for a dilute system, but in the case of the earth it would appear more reasonable not to limit the number of particles in each cube.

If in fact we allow an unlimited number of particles to occupy each cube, then we are dealing with Bose-Einstein statistics. Finding $\bar{n}_j$ in the usual statistical mechanical fashion (Morse, 1969, p. 326) yields

$$\bar{n}(r_j) = \frac{1}{e^{\alpha_1 + \alpha_2 r_j^2} - 1}$$

using the mass and moment of inertia as the constraints so that

$$\bar{\rho}(r) = \frac{\frac{m}{V}}{e^{\alpha_1 + \alpha_2 r^2} - 1} \qquad (6.1)$$

Here we have a problem. There are more unknowns than constraints. If we knew how to

choose $\frac{m}{V}$, then we could find α_1 and α_2 from the constraints of mass and moment of inertia. Unfortunately, we have no clear guidance in this matter.

There is a way, however, to neatly sidestep the problem. We introduce a third piece of information: we assume we know $\bar{\rho}(a_E)$, the value of the density at the earth's surface. Using this information in eq. (6.1) yields

$$\bar{\rho}(r) = \frac{(e^{\alpha_1} e^{\alpha_2 a_E^2} - 1)\bar{\rho}(a_E)}{e^{\alpha_1} e^{\alpha_2 r^2} - 1}$$

and we use our knowledge of $\bar{M}_E$ and $\bar{C}$ to find the two multipliers. Taking $\bar{\rho}(a_E)$ as 2.84 gm/cm^3 to agree with the surface density of rocks (Stacey, 1969, p. 104), we find by numerical integration

$$\bar{\rho}(r) = \frac{5.517}{0.44e^{1.48r^2/a_E^2} + 0.0097} \quad \text{gm/cm}^3 \qquad (6.2)$$

This distribution is plotted in Fig. 2, along with Bullen's (1975) optimum distribution. This Bose-Einstein distribution is almost indistinguishable from the Maxwell-Boltzmann distribution of Figure 1 and the distinction between using the two is academic in this instance, at least.

So we could use Bose-Einstein statistics in the information theory approach to the earth's density distribution. While it is probably conceptually superior to Maxwell-Boltzmann statistics, it is also a little more complicated mathematically, as a comparison of eq.s (6.1) and (6.2) shows. The Maxwell-Boltzmann case in future investigations should probably be investigated first, being simpler.

Another question arises about the information theory approach. Suppose we are again discussing the example of the earth's mass, where we have N masses to choose from

and know that $\Sigma P_i M_i = \overline{M}_E$. The information theory probability distribution is $\exp(\alpha_0 - 1 + \alpha_1 M_i)$ as found earlier. Yet the experimental probability distribution is probably a Gaussian with the form const $\cdot \exp(-(M_i - \overline{M}_E)^2/\sigma^2)$ rather than the simple exponential distribution of information theory, with both $\overline{M}_E$ and σ being known. The same may be said for the moment of inertia. How do we handle this?

It is actually rather easy to obtain a Gaussian distribution from the information theory approach. Assuming that we know

$$-\sum_i P_i (M_i - \overline{M}_E)^2 = -\sigma^2$$

we differentiate

$$-\sum_i P_i \ln P_i + \alpha_0 \sum_i P_i - \alpha_1 \Sigma P_i (M_i - \overline{M}_E)^2$$

with respect to each P_i, giving

$$P_i = e^{\alpha_0 - 1} e^{-\alpha_1 (M_i - \overline{M}_E)^2}$$

which is the desired Gaussian. Presumably finding $\bar{n}(r_j)$ using a Gaussian is difficult, but the question is probably moot. As long as we deal with a narrow range of values of mass near $\overline{M}_E$, it would seem to matter little whether we use exponential or Gaussian distributions.

As we have seen, the information theory approach does have its problems; but the strengths of the approach are many. Its philosophical basis is satisfying: there is no unwarranted weighting of possible answers. It is rational and objective: everyone using the approach will obtain the same answers, given the same data. It gives the "best" answer on the basis of very little data. The mathematics of the theory is standard: that of statistical mechanics. Observational errors have a natural place in the approach. It provides an alternative to extensive modeling. Finally, its generality gives it great power: it can be applied equally well to the statistics of a gas or the interior of the earth. The information theory approach should find broad use in solid earth geophysics.

ACKNOWLEDGMENTS

I wish to thank David F. Smith and Michael A. Graber for many helpful discussions. I thank Professor H. P. Edmundson for pointing out that MI is not dimensionless. The support of a NAS-NRC Resident Research Associateship is gratefully acknowledged.

REFERENCES

Alterman, Z., H. Jarosch and C. L. Pekeris, (1959), Oscillations of the earth, Proc. Roy. Soc. London, 252A, 80-95.

Anderson, D. L., (1975), Structure of the mantle, Rev. Geophys. and Sp. Phys., 13, [3], 300-302.

Backus, G. E., and J. F. Gilbert, (1967), Numerical applications of a formalism for geophysical inverse problems, Geophys. J. R. astr. Soc., 13, 247-276.

Backus, G. E., and J. F. Gilbert, (1968), The resolving power of gross earth data, Geophys. J. R. astr. Soc., 16, 169-205.

Baierlein, R., (1971), Atoms and Information Theory, W. H. Freeman, San Francisco.

Batty, M., (1974), A comment on the paper "A comparison of the Shannon and Kullback information measures", J. Stat. Phys., 11, 523-524.

Bullen, K. E., (1975), The Earth's Density, Chapman and Young, London.

Burg, J. P., (1972), The relationship between maximum entropy spectra and maximum likelihood spectra, Geophysics, 37, 375-376.

Cox, R. T., (1946), Probability, frequency, and reasonable expectation, Am. J. Phys., 14, 1-13.

Cox, R. T., (1961), The Algebra of Probable Inference, Johns Hopkins Press, Baltimore.

Engdahl, E. R., A. M. Dziewonski, and E. A. Flinn, (1975), Structure of the earth's core, Rev. Geophys. Sp. Phys., 13, [3], 302-303.

Gilbert, J. F., A. Dziewonski, and J. Brune, (1973), An informative solution to a seismological inverse problem, Proc. Nat. Acad. Sci. USA, 70, 1410-1413.

Graber, M. A., (1976), Polar motion spectra based upon Doppler, IPMS and BIH data, Geophys. J. Roy. astr. Soc., 46, 75-85.

Graber, M. A., (November 1977), An information theory approach to the density of the earth, NASA Technical Memorandum 78034.

Hobson, A., and B.-K. Cheng, (1973), A comparison of the Shannon and Kullback information measures, J. Stat. Phys., 7, 301-310.

Jaynes, E. T., (1957), Information theory and statistical mechanics, Phys. Rev., 106, Ser. 2, 620-630, and 108, Ser. 2, 171-190.

Jaynes, E. T., (1963), Information theory and statistical mechanics, in Statistical Physics, pp. 181-218, K. W. Ford, editor, W. A. Benjamin, New York.

Jaynes, E. T., (1968), Prior probabilities, IEEE Trans. on Systems Sci. and Cybernetics, SSC4, 227-241.

Jordan, T. H., and J. N. Franklin, (1971), Optimal solutions to a linear inverse problem in geophysics, Proc. Nat. Acad. Sci. USA, 68, 291-293.

Katz, A., (1967), Principles of Statistical Mechanics, W. H. Freeman, San Francisco.

Morse, P. M., (1969), Thermal Physics, second edition, W. A. Benjamin, New York.

Parker, R. L., (1977a), Linear inference and underparameterized models, Rev. Geophys. Sp. Phys., 15, 446-456.

Parker, R. L., (1977b), Understanding inverse theory, Ann. Rev. Earth Planet. Sci., 5, 35-64.

Press, F., (1968a), Earth models obtained by Monte Carlo inversion, J. Geophys. Res., 73, 5223-5234.

Press, F., (1968b), Density distribution in earth, Science, 160, 1218-1221.

Richards, P. G., (1975), Theoretical seismology, Rev. Geophys. Sp. Phys., 13, [3], 295-298.

Rowlinson, J. S., (1970), Probability, information and entropy, Nature, 225, 1196-1198.

Shannon, C. E., (1948), A mathematical theory of communication, Bell System Tech. J., 27, 379-423, and 27, 623-656. (Also, reprinted in The Mathematical Theory of Communication by C. E. Shannon and W. W. Weaver, University of Illinois Press, Urbana, Illinois, 1949.)

Smylie, D. E., G. K. Clarke, and T. J. Ulrych, (1973), Analysis of irregularities in the earth's rotation, in Methods in Computational Physics, B. A. Bolt, editor, 13, pp. 391-430, Academic Press, New York.

Stacey, F. D., (1969), Physics of the Earth, first edition, Wiley, New York.

Tribus, M., (1961), Thermostatics and Thermodynamics, Van Nostrand, New York.

Tribus, M., and E. C. McIrvine, (1971), Energy and information, Sci. Am., 225, [3], 179-188.

Tribus, M., and R. Rossi, (1973), On the Kullback information measure as a basis for information theory: comments on a propcsal by Hobson and Chang, J. Stat. Phys., 9, 331-338.

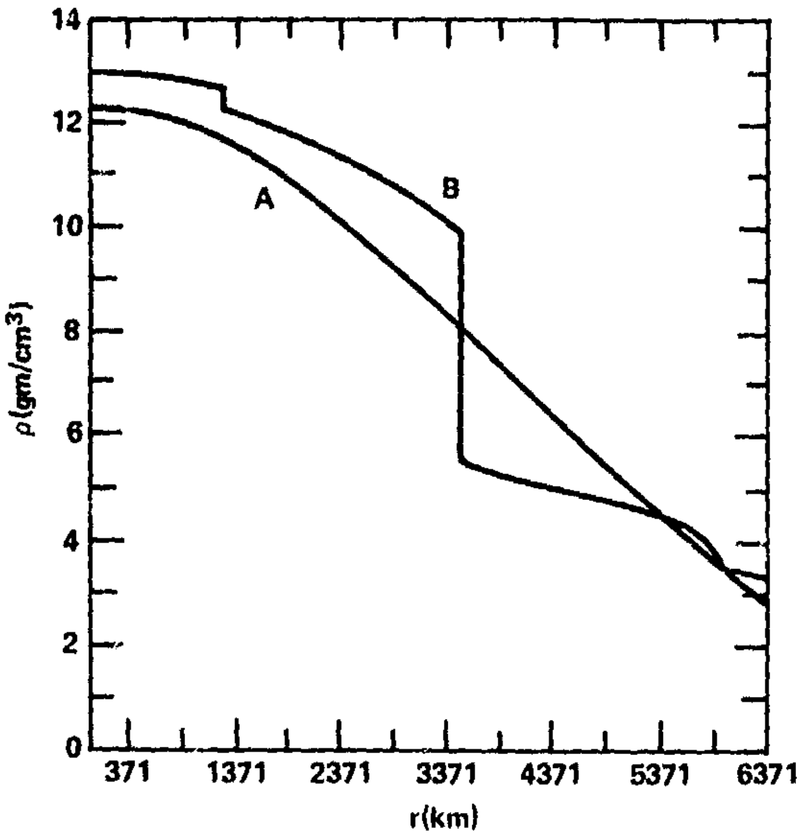

Figure 1. The information theory density distribution using Maxwell-Boltzmann statistics (curve A) and the optimum density distribution of Bullen (1975) (curve B) are shown as a function of radial distance r.

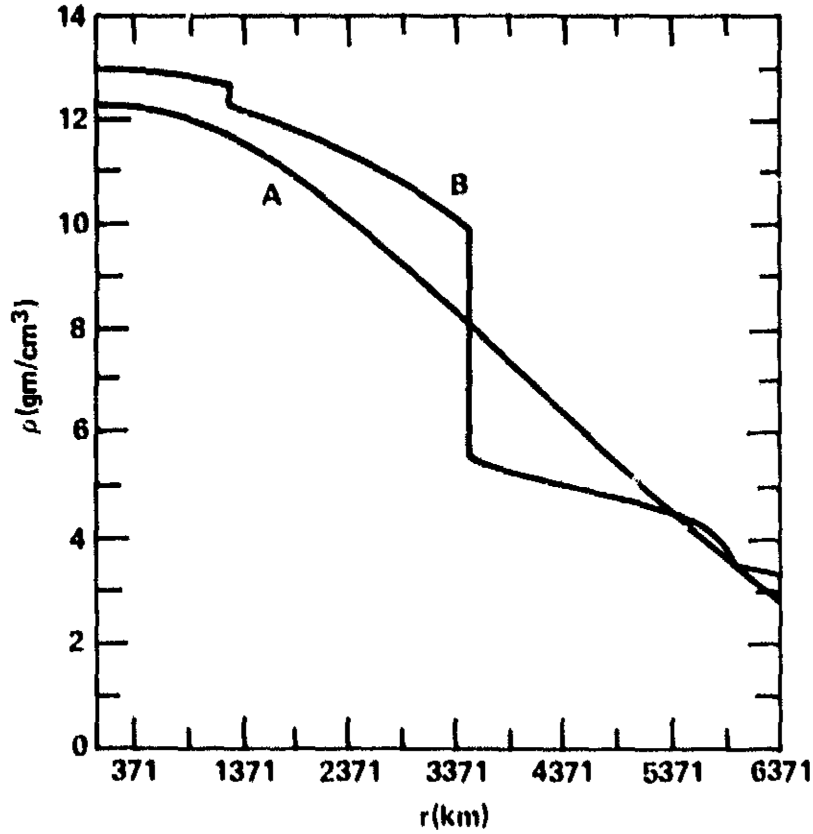

Figure 2. The information theory density distribution using Bose-Einstein statistics (curve A) and the optimum density distribution of Bullen (1975) (curve B) are shown as a function of radial distance r.

BIBLIOGRAPHIC DATA SHEET

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
TM 78088		

4. Title and Subtitle	5. Report Date
Information Theory and the Earth's Density Distribution	February 1978
	6. Performing Organization Code
7. Author(s) David Parry Rubincam	8. Performing Organization Report No.
9. Performing Organization Name and Address Goddard Space Flight Center Geodynamics Branch, Code 921 Greenbelt, MD 20771	10. Work Unit No.
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address Same as above	13. Type of Report and Period Covered
	14. Sponsoring Agency Code

15. Supplementary Notes

Submitted to Geophysical Journal

16. Abstract

The present paper argues for using the information theory approach of Jaynes(1957) as an inference technique in solid earth geophysics. A spherically symmetric density distribution is derived as an example of the method. A simple model of the earth plus knowledge of its mass and moment of inertia leads to a density distribution which is surprisingly close to the optimum distribution of Bullen(1975). Future directions for the information theory approach in solid earth geophysics as well as its strengths and weaknesses are discussed.

17. Key Words (Selected by Author(s))	18. Distribution Statement
Information theory, density distribution	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	23	

*For sale by the National Technical Information Service, Springfield, Virginia 22151.

GSFC 25-44 (10/77)